

AD-A079 243

ARMY PERSONNEL RESEARCH OFFICE WASHINGTON DC
MODERATOR VARIABLES: A REVIEW OF THE LITERATURE.(U)
JUN 65 P A BANAS

F/G 5/10

UNCLASSIFIED

APRO-RM-65-4

NL

| OF |

AD
A079243L



END

DATE
FILMED

2-80

DDC

LEVEL 4

2

ADA 079243

Research Memorandum 65-4

**MODERATOR VARIABLES: A REVIEW
OF THE LITERATURE**

June 1965

DDC
RECEIVED
DEC 19 1979
RECEIVED

DDC FILE COPY

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited



U. S. ARMY PERSONNEL RESEARCH OFFICE

79 12 18 205

12 27

Army Project Number
2J024701A722

NCO Leaders a-17

9 Research Memorandum 65-4

6 MODERATOR VARIABLES: A REVIEW OF THE LITERATURE.

10 by Paul Anthony Banas

Francis F. Medland, Task Leader

14 APPRO-RM-65-4

Submitted by:
Leonard V. Gordon
Chief, Behavioral Evaluation
Research Laboratory

Approved by:
J. E. Uhlaner
Director of
Laboratories

11 June 1965

Research Memorandums are informal reports on technical research problems. Limited distribution is made, primarily to personnel engaged in research for the U. S. Army Personnel Research Office.

040650 Gm

FOREWORD

The literature review constituting the present Research Memorandum is the introductory portion of a dissertation submitted to the graduate faculty of the University of Minnesota in March 1964 in partial fulfillment of the requirements for the PhD degree. The author, Dr. Banas, completed the review and the research portion of his dissertation while a member of the staff of the U. S. Army Personnel Research Office. The research study proper was reported in USAPRO Research Memorandum 64-10, Empirical Transsituational Moderators, November 1964.

The review has been reproduced exactly as prepared by the author, primarily for use of the staff of the U. S. Army Personnel Research Office. Mention of commercial test organizations and their products by name was essential to precision in reporting research findings and does not constitute indorsement of any testing instrument.

| | |
|--------------------|--|
| Accession For | |
| NTIS GRA&I | <input checked="checked" type="checkbox"/> |
| DDC TAB | <input type="checkbox"/> |
| Unannounced | <input type="checkbox"/> |
| Justification | |
| By | |
| Distribution/ | |
| Availability Codes | |
| Dist | Availand/or special |
| A | |

A REVIEW OF MODERATOR RESEARCH¹

A comprehensive survey by Ghiselli (1955) of both published and unpublished studies dramatically points up the predictive plateau which has been reached with the classical validation model in industrial test and selection research. The average validities ranged in the 30's and lower 40's; an upper limit of .50 was evident. This represents the upper limit for tests predicted by Clark Hull (1928) 35 years ago. Also, in reviewing the progress of academic test and selection research, England (1960) points out that, "Although considerable research effort has been expended, college grades are not predicted more accurately today than was the case 10 or 15 years ago." This state of affairs in industrial and educational test and selection research has led to dissatisfaction with the classical prediction and psychometric models and to the search for more appropriate complex models.

Among the most ardent proponents of modification of the classical models have been Edwin Ghiselli and Marvin Dunnette. Ghiselli (1963a), in his presentation of the Donald G. Paterson Lecture at the University of Minnesota, cited his objection to classical psychometric theory. He stated, "I fear that classical psychometric theory has so dominated our thinking that we are blind to other possible explanations of differences among individuals in their predictability--and other possible explanations there indeed are." He discussed the rudiments of a psychometric theory which allows for the differential psychological structure of individuals. In other words, a given or basic underlying trait need not have the same degree of importance for all individuals in determining a particular behavior. In contrast with classical psychometric theory, it is assumed that individuals are not equally reliable or predictable. Therefore, subgrouping on the basis of differential error of measurement or differential error of prediction should lead to differential patterns of validity.

In rejecting the classical validation model, Dunnette (1963) states, "The first step toward enhancing the prediction of human behavior must, therefore, be to accept the complexities of human behavior and adopt a prediction model which is appropriately complex." He recommends a modification of the classical validation paradigm suggested by Guetzkow and Forehand (1961), which he thinks provides a richer schematization for prediction research, and which offers important implications for the direction of future research. The most important implication is that differential patterns of validity for various subgroups must be recognized and that research should be devoted to the definition of homogeneous subsets of individuals within which appropriate prediction equations may be developed and cross-validated. Both Ghiselli and Dunnette are in accord in their emphasis of differential patterns of validity for various subgroups.

¹This work was completed while at the U.S. Army Personnel Research Office, Washington, D.C., and was submitted to the faculty of the University of Minnesota in partial fulfillment of the requirements for the Ph. D., March 1964. The author wishes to express his gratitude to H. P. Longstaff, M. D. Dunnette, L. L. Lofquist, and E. E. Ghiselli for their helpful advice and encouragement, and the U.S. Army Personnel Research Office for partial support of this work. The author is presently associated with Human Sciences Research, Inc., McLean, Virginia.

To a great extent, the emphasis on differential patterns of validity stems from recent studies in the literature which show that subgrouping analysis has, indeed, led to differential patterns of validity and enhancement of prediction. Evidence presented by Kahneman and Ghiselli (1962) and DeSoto and Kuethe (1960), bearing on the inappropriateness of the linear homoscadastic model and the abuse of the Pearson product-moment correlation is also pertinent. They emphasize the importance of considering the interaction between variables and the isolation of homogeneous subgroups.

Two general approaches to the isolation of homogeneous subgroups which consider the interaction between variables are the pattern-analytic techniques and the identification of moderator variables. The present review is concerned with the identification of moderator variables.

DEFINITION OF A MODERATOR VARIABLE

The variables which have been used to define the subgroups differing in patterns of validity have been referred to as moderator, predictability, modifier, referent, and population control variables. The definitions given to these terms have been as follows:

moderator variable (Saunders, 1956)

A continuous variable which systematically affects the predictive effectiveness of another psychological measure, that is the weight given the measure is a function of the individual's score on this continuous variable.

population control variable (Gaylord and Carroll, 1948)

A variable which identifies subpopulations in which the application of a multiple regression equation, optimum for the entire population, is inappropriate.

modifier variable (Grooms and Endler, 1960)

An independent variable which when dichotomized or trichotomized leads to differential subgroup relationships between a predictor variable and a criterion variable.

predictability variable (Ghiselli, 1956, 1960a, 1960b)

A variable or empirical scale which is positively correlated with the D (difference) score, regardless of sign, between the standard predictor and standard criterion scores. It classifies individuals on a better than chance basis in terms of the extent to which the scores they earn on a predictor are related to their scores on a criterion variable.

referent variable (Toops, 1948, 1959)

A variable, status or environmental, which modifies the meaning--the causal potency--of other variables. All such referent variables imply interactions, that is, the weight of a predictor variable is not a constant, as implied in ordinary regression equations, but rather is a mathematical function of a referent variable.

The common characteristics are implied in the definitions of these terms. First, they all refer to situations in which the relationship between two or more given variables is found to vary as a function of changes in the value or character of one or more other variables. Thus, all the definitions imply some form of interaction between variables. Second, the definitions all imply that subgroups, homogeneous in some quantitative or qualitative respect, may be isolated which show differential patterns of validity.

Since the differences between the uses of the terms cited above are vague and confused in the minds of most psychologists, it is proposed that moderator variable be accepted as the general term to refer to all variables, quantitative or qualitative, which improve the usefulness of a predictor by isolating subgroups of individuals for whom a predictor or set of regression weights are especially appropriate. Therefore, under this more general definition, an aptitude, attitude, personality, interest, status, situational variable or empirically derived scale may function as a moderator.

Although the term moderator is recent in origin (Saunders, 1956), in actuality, the recognition of the joint functional relationships between variables and the enhancement of prediction by subgrouping are not new. Methods capable of handling the joint functional relationships between variables have existed for at least 30 years (Court, 1930) and as early as 1926, if not earlier, evidence existed pertaining to the differential patterns of validity, i.e., differential predictability of various subgroups (Scates, 1926).

The fact that moderator effects do occur has been substantiated by a wealth of evidence found in studies in industrial, educational, and military settings. Evidence, however, also exists to temper the enthusiasm of individuals who view this concept with extreme optimism. A review of the literature will indicate the types of moderator variables that have been identified. It will also serve to lead to a discussion of questions relevant to moderator research.

REVIEW OF THE LITERATURE

[The objectives of this review are (1) to report the research pertaining to the development of the moderator concept, (2) to indicate the wide breadth of potential application of the moderator concept, and (3) to identify the problem areas in moderator research.] Studies reported by psychologists in educational, industrial, and military settings which are representative of both the empirical and rational approaches to the identification of moderator variables have been selected for inclusion in this review.

A. Empirical Approach to the Identification of Moderators

Intra-individual variability has been a matter of sporadic interest to psychologists since the time of Hull (1927, 1928). However, as Fiske and Rice (1955) have indicated in their summary of the literature, it is readily observed but has proved to be only of limited predictive value.

In an effort to assess the possible usefulness of intra-individual variability as a moderator, Berdie (1961) determined a variability index on the basis of variability of subscores on a 100 item algebra test which was subsequently used as a predictor of first year grade point average and mathematics course grades. The hypothesis was that first year engineering students whose variability of behavior was greater while taking a test used for predictive purposes would be less predictable than persons whose behavior was less variable. The results of the analysis of the data tended to support the hypothesis. Most of the differences were in the expected direction, but only some were statistically significant. A few of the differences, although not statistically significant, were in the opposite direction. In summary, the results of this study reveal that a consistent and meaningful measure of intra-individual variability can be obtained from responses on a psychological test. Differences in intra-individual variability are related to a limited extent to predictability, but the method of determining intra-individual variability is important. For instance, the variability index derived from a form with test items arranged roughly in order of difficulty was better than an index derived from a form with items cyclical in order of difficulty. Although the relationship between predictability and intra-individual variability seems to be rather complex, the results of this study indicate that further research should be undertaken in this area.

The study of intra-individual variability by Berdie (1961) can be viewed as an attempt to ascertain whether a predictor test can also serve as its own predictability test, i.e., moderator. Ghiselli (1960a) undertook an investigation with this objective in mind. Item analyses were performed on the responses of college students to a self-description inventory to provide moderator keys for an "intelligence" scale and "initiative" scale. These scales were derived from the inventory and used as predictors. Scores on a questionnaire designed to measure "sociability" were used as the criterion. The results of this study provide substantiation for the use of empirically derived moderators. It also demonstrates that an inventory can be used both to predict criterion variables and to indicate the accuracy with which the prediction can be made in individual cases.

Using other scales derived from this self-description inventory as predictors, and supervisory ratings as criteria, Ghiselli (1963b) provided further evidence to substantiate the value of empirically derived moderator scales in enhancing validity. More predictable and less predictable subgroups were differentiated from three different samples--factory worker, executives and foremen. He showed that those individuals for whom a test has greater validity can be systematically differentiated from those for whom it has a lesser degree of validity.

The empirical derivation of moderator scales by Ghiselli (1960a, 1963b) was based on the selection of items which correlate most highly with the absolute differences between the standard scores on the predictor and the standard scores on the criterion. Ghiselli (1960b) has also used this technique to investigate the possibility of developing scales that would give objective indication of which of two tests predict better for an individual. Using two different samples, a

college sample and an industrial sample, different predictors and different criteria, he showed that moderator scales could be empirically developed through item analyses which would indicate in an objective way which of two tests give the better prediction for a given individual. Scales of this type could be invaluable to the psychometrician in selection and counseling situations if found to be practically applicable.

No studies were found in the literature where the investigator tried to develop a moderator by item analysis in which the algebraic differences between the standard predictor scores and the standard criterion scores were used as the dependent variable. A "deviate technique", where the item responses are correlated with the actual-minus-predicted criterion measures has been used by Neidt and Malloy (1954) and England (1960). Instead of subgrouping on the basis of the scales, these investigators used the scales as a predictor in a multiple regression equation. Equivocal results were obtained.

In addition to a consideration of the deviation from the Line of Relations (Ghiselli, 1956, 1960a, 1960b, 1963b), two other sources of deviation have been employed to develop empirical moderator scales. These are the statistical infrequency of response and response inconsistency.

Johnson and Roy (1958) developed two moderator scales for the detection of irrelevant or careless responses to a personality inventory. The purpose was to refine the sample to be used for item analysis by eliminating the individuals who gave evidence of responding in a careless or random manner. The "frequency" scale consisted of items to which nearly all subjects respond in the same way. The "consistency" scale included pairs of "Yes-No" items (presented separately), so stated, that response to the first member of the pair dictated the logically consistent response to the other member. Comparison of the validity coefficients for cross-validation samples from the unrefined and refined item analysis samples was made. The results indicated that stability of item validity coefficients can be improved considerably through the use of a carelessness moderator. The carelessness scales were also found to be highly reliable.

In this case, the authors point out that the carelessness moderator, like a distortion scale, has the characteristics of a suppressor variable. Therefore, instead of rejecting cases on the basis of carelessness scores, the scales could be used to correct inventory scores to an extent indicated by the carelessness score. The K scale of the Minnesota Multiphasic Inventory is a prime example of a scale used essentially as a correction factor to sharpen the discriminatory power of the clinical variables measured by the inventory (Meehl and Hathaway, 1946).

Verification, validity, or carelessness moderator scales based on the statistical infrequency or frequency of responses have been developed for the Strong Vocational Interest Blank (Filbeck and Callis, 1961), the Kuder Personal Preference Record (Kuder, 1956), the Minnesota Vocational Inventory

(Campbell and Trockman, 1963) as well as the Minnesota Multiphasic Inventory (Meehl and Hathaway, 1946). Usually the rationale is that anyone who selects a substantial number of low frequency items is being careless, does not understand the directions, or is deliberately trying to invalidate the results. Inclusion of these invalid cases in an experimental investigation of an instrument could lead to spurious results.

Banas (1964) explored the possibility of identifying transsituational moderators. These are moderators that could function to identify more and less predictable subgroups when applied in different situations. This may mean different predictors, different criteria, different samples, or any combination thereof. Previous research seems to indicate that moderators do not show transsituational characteristics, but are specific to a particular situation.

The specific questions related to transsituational moderators asked in this study were: (1) Can a moderator be found that identifies more and less predictable subgroups for different occupational groups when different aptitudes related to performance are used as predictors for each of the groups? (2) Can a moderator be found that identifies more and less predictable subgroups for different occupational groups when the same aptitude related to performance is used as a predictor for each of the groups? Both a rational and empirical approach were used.

The samples consisted of handicapped and nonhandicapped individuals who participated in a study of work adjustment being conducted by the Regional Vocational Rehabilitation Research Institute at the Industrial Relations Center, University of Minnesota. Clerical, nonskilled, and skilled workers were considered. Each of the samples consisted of approximately 50% handicapped persons. The samples were divided into an experimental and cross-validation sample for the empirical analysis. The total sample was used for the rational approach.

The predictors consisted of scales of the GATB; rational and empirical predictors were selected and evaluated. The effectiveness of the moderators was determined for both types of predictors. Supervisory ratings were used as a criterion.

The scales of the MMPE, scales from the SVIB, and the G scale of the GATB were examined as potential moderators in the empirical analysis. Age and disability were examined as rational moderators. The two major hypotheses tested for the rational moderators were: (1) There is no difference in predictability of younger and older workers when an aptitude predictor related to performance for an occupational group is used as a predictor; (2) There is no difference in the predictability of handicapped and nonhandicapped workers when an aptitude predictor related to performance for an occupational group is used as a predictor.

The empirical selection of moderators was based on Ghiselli's difference score technique. This method consists of computing difference scores between the standard scores on the predictor and the criterion, and determining the effectiveness of the scale as a moderator by examining the correlation of the scale with these absolute difference scores. The higher the scale correlates with the absolute difference scores, the more effective the scale is as a moderator.

A modified Wherry-Doolittle test selection technique was used to develop composite specific and transsituational moderators. The specific moderator was developed on each sample individually; the transsituational moderator was developed on the combined samples.

Transsituational moderators were found for the rational predictors, but not for the empirical predictors. One identified more predictable subgroups for different occupational groups when different aptitudes were used as predictors for each of the groups, and the other identified more predictable subgroups for different occupational groups when the same aptitude was used as a predictor for each of the groups. The transsituational moderators were effective for the clerical and skilled samples, but not for the nonskilled sample. These results tend to support the conclusion that transsituational moderators do exist and can be empirically developed.

The rational moderators, age and disability, were found to function differently for these samples. Age was found to be specific, and the research hypothesis that younger workers are more predictable than older workers when aptitudes are used as predictors was rejected. Disability was found to be transsituational, and the research hypothesis that nonhandicapped workers are more predictable than handicapped workers when aptitudes are used as predictors was supported.

In general, the results of this study support the moderator concept. More and less predictable subgroups were identified by both rational and empirical moderators. In particular, these results seem to indicate that (1) empirical transsituational moderators can be found using the method employed in this study, and that (2) nonhandicapped workers seem to be more predictable than handicapped workers when aptitudes are used as predictors.

B. Rational Approach to the Identification of Moderators

Rather than develop moderators through an empirical approach, other researchers have studied interest, personality, attitude, and biographical variables as potential moderators.

The psychologist in the educational setting, for many years, has been interested in the prediction of academic success and the effect of various variables on the relationship between predictors and criteria. Among the earliest studies of the effect of a variable on this relationship was a study by Scates (1926). He found that the grades of college students graduated from certain high schools were more predictable than others. This differential predictability of students according to the high school from which they were graduated received further support in a study of homogeneous subgrouping as a means of improving prediction (Wagner and Strabel, 1935). Furthermore, as in more recent studies by Rundquist (1941) and Abelson (1952), Wagner and Strabel (1935) found that the academic performance of women was more predictable than that of men. Seashore (1961) has summarized a vast number of studies which uniformly show that the grades of women (in both high school and college) are significantly more predictable than those of men.

Marks (1964) presents data which indicate the potential usefulness of undergraduate grades as a moderator. Aptitude test scores (X), undergraduate grades (Y), and graduate grades (Z) were available for 50 graduate students in the College of Business Administration at the University of Rochester. X and Y yielded validity coefficients of .40 and .50 respectively, and multiple R using X and Y was .61. The sample was split into upper and lower halves on the Y score and r_{XZ} was found to be .70 and .20 in the two halves. Z was predicted by $Z = X(A + BY)$, where A is r_{XZ} for the whole sample and B is the standard deviation of the .70 and .20 values. In all instances, Z predicted by the moderator approach had a higher crossed validity than did a multiple-regression equation.

Marks states that, in hindsight, the results make psychological sense. Those students with high undergraduate grades (Y) can be assumed to have been highly motivated (unless they cheated); in such a group the aptitude test should be, and is, a very effective predictor of variance in graduate work. But the students with low undergraduate records could have achieved those grades by being poorly motivated, by being inept, or by some combination of both factors. In such a heterogeneous group, one could not expect that the aptitude score could have much predictive value, and it did not. Marks suggests that these unpredictable or off-quadrant cases are a very fertile ground for future personnel research.

Some investigators have been interested in the effect of personal adjustment on the relationship between scholastic aptitude and academic criteria. The common-sense notion that well-adjusted students should be more predictable than maladjusted students was supported by Stagner's (1933) finding that students scoring high on the Bernreuter Personality Inventory were more predictable with the American Council on Education Test than those scoring low. More recently, Hoyt and Norman (1954) investigated the effect of personal adjustment on the relationship between predictors and criteria. Using two or more T scores above

70 on the scales of the Minnesota Multiphasic Inventory as a criterion of maladjustment and no T scores above 60 as a criterion of normality, they concluded that the accuracy of prediction of grades was a function of the state of adjustment of the student. Therefore, it would appear that separate regression equations for adjusted and maladjusted students should be used. The use of separate regression equations, however, was not supported by Gallese (1958) in his study of normal and maladjusted students as defined by Minnesota Multiphasic profiles.

In a replication of the Hoyt and Norman (1954) study, Anderson and Spencer (1963) obtained negative results. They concluded that personal adjustment, as defined by Hoyt and Norman, does not influence the prediction of achievement of arts college or engineering freshman students. It would appear that the definition of personal adjustment of Hoyt and Norman (1954), and that of Gallese (1958), may be inadequate. Redefinition of adjustment in terms of more meaningful profile analysis, or consideration of interaction between personality scales might be more fruitful.

Grooms and Endler (1960) and Malnig (1959) have studied the effect of anxiety on the relationship between scholastic aptitude and achievement. Grooms and Endler used the Mandler-Sarason Test Anxiety Questionnaire and Malnig used the Taylor Manifest Anxiety Scale as a measure of anxiety. Both reported that test anxiety serves as a moderator which can be used to enhance the predictability of grades from aptitude test scores. However, Malnig found that low anxiety students were more predictable, and Grooms and Endler found that high anxiety students were more predictable. It would certainly appear that this discrepancy in findings can be attributed to the use of two different measures of anxiety.

One of the promising newer approaches to the problem of prediction of academic success has been the attempt to predict academic success differentially for the different curricula. Horst (1954, 1955) has developed elaborate statistical procedures for selecting from a large battery, a subset of tests which will predict differential success most efficiently in a large number of curricula. However, in a recent study of the relationship between preadmission variables and success in college, Horst (1959) reports data that seem to raise questions as to the value of subgrouping on the basis of curricula for the prediction of academic success. High school grades in relevant subjects were used as predictors of future success in related college curricula. In a similar attempt to predict academic success differentially for different curricula, Eells (1961) also obtained negative results. His predictors were aptitude, reading, and achievement tests and his samples were liberal arts, engineering, and commerce subgroups. The evidence appears to warrant further investigation for the specific situation before substituting differential prediction for overall predication for the prediction of grades.

In general, personality variables have not proved to be very useful as predictors of academic success, whether used individually or in combination with other variables. Recently, however, some investigators have reoriented their research to study personality scales as predictors when other variables are used as moderators. Goodstein and Heibrun (1962) studied the prediction of college achievement from the Edwards Personal Preference Schedule when intellectual ability was treated as a moderator variable. The results of this study offer some support for the notion that personality factors are significantly related to academic achievement when the influence of academic ability is controlled by homogeneous subgrouping. However, the nature of the relationship is dependent upon the general ability level of the group being studied. That is, the personality scales improved prediction, but only for the middle ability group. One possible general interpretation of these results given by the authors is that personality factors are most important in determining academic achievement of the average ability college male, and that success of relatively bright and dull males is more determined by intellectual factors. This finding is not supported by evidence cited by Gordon (1963a, 1963b) related to the moderating effect of scholastic aptitude upon the validity of the scales of the Gordon Personal Profile and Gordon Personal Inventory. He interprets the results as supporting the hypothesis that certain personality traits have greater predictive efficiency at high aptitude levels. His rationale is that systematic variation in validity with aptitude level will tend to occur in circumstances where a given level of aptitude is requisite for passing a particular academic program. In such instances, for individuals within the submarginal aptitude range, non-aptitude measures will generally show no relationship to achievement, that is, high status on relevant interest or personality scales will not compensate for inadequate aptitude. However, beyond the point of marginal aptitude, non-cognitive factors will tend to play a role in increasing importance. The predictions made by Gordon are not supported by the work of Holland (1960) and Holland and Astin (1962), who studies National Merit finalists. In Holland's studies of the predictive effectiveness of many scales from the Strong Vocational Interest Blank, California Psychological Inventory, 16 Personality Factor Questionnaire, and a number of other instruments, he concluded that for this very homogeneous high ability group the findings were consistent with those reported in previous research, which used similar variables on less able students. Further evidence of this lack of predictive effectiveness of personality scales for a homogeneous high ability group is presented by Uhlinger and Stephens (1960) in their study of the relation of achievement motivation to academic achievement. No significant relationships were found between actual achievement and need for achievement, as measured by three different scales. The intercorrelations between the three different measures of need for achievement were close to zero.

The lack of agreement in the studies cited above is quite descriptive of the present state of personality research. This failure to provide clear-cut evidence either for or against the notion of the moderator variable is evidently

due to the lack of comparability of groups studied, or the lack of sensitivity and validity of the measurement devices employed. Fiske (1963) has stated this argument quite well. "The measurement of personality variables is in a highly unsatisfactory state at the present time. It is poor not only in absolute terms, but also relative to the measurement of abilities, aptitudes, and achievement. The typical personality test has low internal consistency, considerable instability over time, and very little predictive value. Perhaps most serious of all is the limited degree of convergent validation: Measures purporting to assess the same variable typically show low inter-correlations."

Although interest tests differentiate better among curricula groups than other kinds of tests, and can predict the curriculum from which a student will graduate better than aptitude, achievement, or personality tests (Berdie, 1955), interests measures have proved to be no more effective than personality tests when it comes to predicting academic success. Certain common sense notions such as, "grades achieved in academic subjects should be commensurate with interest in the subject", have not been supported by research findings even when the level of ability is controlled. This failure to obtain high relationships between level of interest and level of achievement led Frederiksen and Melville (1954) to investigate the effect of level of compulsiveness on the predictive efficiency of interest scales. They argued that if they could find some way of identifying compulsive and non-compulsive students, then they could enhance the prediction of interest measures. They identified students whose speed of comprehension score was higher in relation to their scholastic aptitude score as non-compulsive, and students whose speed of comprehension score was lower in relation to their scholastic aptitude score as compulsive. A second method of dividing students into compulsive and non-compulsive subgroups was based on the accountant score of the Strong Vocational Interest Bland. Those above the mean of the group constituted the compulsives and those below the mean the non-compulsives. Ten scales previously found to be relevant predictors for engineering students were used as predictors.

The hypothesis that the correlation between interest measures and achievement is higher for students who are relatively free from compulsive tendencies tended to be supported for both measures of compulsiveness. These two measures did not correlate with each other, and when combined, differential predictability was even greater for those subgroups identified as compulsive on both scales. However, in a replication of this study by Frederiksen and Gilbert (1960), the finding that non-compulsive students were more predictable with interest scales held up for only one measure of compulsiveness and then with only four of the ten interest scales. An interpretation suggested by the data of Frederiksen and Melville (1954) that compulsiveness, as measured in their study, showed potential as a transsituational moderator, that is it enhances the prediction of a number of relevant predictors for a particular sample and criterion, is only partially supported by the results of Frederiksen and Gilbert (1960).

Using the Frederiksen and Melville (1954) data, Saunders (1956), who coined the term moderator variable, assessed his method for dealing with the moderator as a continuous variable rather than creating differentially predictable subgroups. This method involves a type of multivariate curvilinear regression in which beta weights, instead of being constant, are linear functions of the moderator variables. He compared his moderated regression method with the multiple correlation method, where the moderator was also treated as a predictor. In five of the ten comparisons, the moderated regression resulted in a significant increase in prediction over the multiple correlation method. However, in nine of the ten comparisons between absolute correlations for the non-compulsive subgroups and the linear and moderated multiple correlations for the total group, the subgroup correlations were higher than either of the methods using the total group. It would seem more advisable to use subgrouping methods for enhancing prediction rather than Saunders' moderated regression method unless circumstances preclude the subgrouping approach (Johnson, 1960).

In a cross-validation analysis, Saunders (1956) compared the predictive effectiveness of the linear and moderated regression methods. He considered only those scales of the Strong Vocational Interest Blank that had been studied previously by Frederiksen and Melville (1954) and on which he had done prior analysis. The results supported the higher predictive effectiveness of the moderated regression method over the linear multiple regression method. No subgroup correlations were reported for this data.

A unique measure of interest was used as a moderator variable in a study by Holtzman, Brown, and Farquhar (1954). They informed students that individual interpretations of their scores on the Survey of Study Habits and Attitudes were available by contacting one of the above authors. When the correlations between the Survey of Study Habits and Attitudes scores and first semester grades were recomputed for subsamples of men and women who later asked for interpretations as opposed to those who did not inquire, the correlations for both the men and women who made inquiries were significantly higher. This suggests that other inventories may have higher validity when used for students who are seeking help with academic, personal, or vocational problems.

As reported above (Scates, 1926; Wagner and Strabel, 1935), the high school from which a student is graduated appears to affect the predictability of his grades with scholastic aptitude tests. McArthur (1954) presents results which indicate that the school attended also affects the relationship between interest measures and occupational choice. His hypothesis that interests would matter less for private school boys, and, therefore, prediction would be poorer than for public school boys, was supported in a follow-up of Strong Vocational Interest Blanks administered fourteen years previously. In this same vein, Dunnette (1962) reports the results of an unpublished study by Owens who recognizes two other subcultures. Differential patterns of validity were obtained on a biographical questionnaire for predicting scholastic achievement of rural and urban students.

The recent interest displayed by industrial psychologists in the differentiation of individuals in terms of their predictability stems to a great extent from the stimulating work of Dr. Edwin E. Ghiselli (1956, 1960a, 1960b, 1963a, 1963b). Using various occupational groups, and both an empirical and rational approach, he has identified scales and variables which functioned as effective moderators.

In a study of taxi drivers (Ghiselli, 1956), an occupational level inventory was found to be an effective moderator for a tapping and dotting test. For the one third of the individuals in the cross-validation sample whose scores on the occupational level inventory indicated that their job performance should be quite predictable from the tapping and dotting test, the validity coefficient was found to be .66, whereas the correlation was only .26 for the total group. The results of this study suggest that the inclusion of individuals whose goals are appropriate and individuals whose goals are inappropriate in a validation study may mask the predictive power of the aptitude being evaluated. In another study of taxi drivers, Ghiselli (1960b) investigated the possibility of selecting which of two tests would be a better predictor for an individual driver. He found that a combination of age and education would discriminate between those individuals for whom a spatial test (discrimination of distance) or a motor test (tapping and dotting) would be a better predictor. Those individuals for whom the spatial test gave better prediction tended to be older and have had less education.

In an investigation of employed men varying occupationally from higher management to unskilled laborers, Ghiselli (1960b) found that an empirical moderator scale was effective for the cross-validation group in discriminating between those individuals for whom a supervisory scale or initiative scale would be a better predictor. He also used this empirical approach to develop moderator scales for samples of factory workers, foreman, and executives (Ghiselli, 1963a). In the case of the factory workers, the predictor was a popularity scale and the criterion was supervisory ratings. For the foreman and executives, the predictor was a supervisory ability scale and the criterion was also supervisory ratings. In all cases, the empirical moderator was found to be effective in enhancing validity when applied to the cross-validation samples.

Using the same procedure he employed to study moderating effects of validity, Ghiselli (1963a) further examined moderating effects on the reliability of measurements. Two parallel forms of a complex reactions test were administered to a sample of semiskilled workers. A combination of age, education, and scores on a tapping test was found to be an effective moderator when applied to the cross-validation sample. The reliability coefficient, the correlation between the two forms of the test, was .92 for the entire group, and progressively increased as the sample was restricted to individuals scoring higher on the moderator combination.

A number of other industrial psychologists have contributed to the literature on homogeneous subgrouping and moderator effects. With respect to job groupings, Dunnette and Krichner (Dunnette, 1958; Dunnette and Kirchner, 1958; Dunnette and Kirchner, 1960) have studied the different patterns of validities obtained when jobs are grouped which are relatively homogeneous in terms of actual responsibilities. Substantially different validities were obtained for subgroups of engineers, salesmen, and clerical employees when grouped according to functional similarities. Wallace and Twichell (1949) found similar results with insurance salesmen. Validities for an aptitude index were found to differ according to whether the agent was commissioned or salaried. This finding of differential validities according to subgroup similarities, as Dunnette (1963) has pointed out, is inherent in the methods of synthetic validity (Balma, et al., 1959) and in the use of the J-coefficient developed by Primoff (1955), which emphasizes the varying predictability of job activities.

The effect of moderator variables on the relationship between personality, attitudinal, situational and criterion variables has been studied by other industrial psychologists. Vroom (1960) studied the moderator effects of two personality variables thought to be relevant in determining an individual's response to participation. The relationship between a person's perception of the amount of influence he had upon the decisions in his work group (psychological participation), and his attitude toward his job was found to be a function of need independence and authoritarianism. The relationship was higher, the higher his measured need independence or the lower his authoritarianism as measured by the California F-Scale. Furthermore, the relationship between psychological participation and supervisory ratings of performance was also increasingly higher, the higher the need independence and the lower the authoritarianism. For the sample of first, second, and third line supervisors, Vroom concluded that authoritarians and persons with weak independence needs are apparently unaffected by the opportunity to participate in decision making. On the other hand, equalitarians and those who have strong independent needs develop more positive attitudes toward their job and greater motivation for effective performance through participation. This evidence indicates that the interaction between participation and personality cannot be ignored; to do this means overestimating the effects of participation on some persons and underestimating the effects of participation on others.

In another study with Mann (Vroom and Mann, 1960), the relationship between the supervisor's authoritarianism and the subordinate's attitude toward work was studied. Differential correlation coefficients were found when groups of drivers, who worked independently of each other, but reported to the same supervisor, were compared with groups of positioners who worked in crews and reported to the same supervisor. For the drivers, authoritarianism of the supervisors correlated positively with the attitudes of the workers; for the positioners, authoritarianism of the supervisors correlated negatively with the attitudes of the workers. Obviously, the results of this study can be interpreted in terms

of moderator effects; however, the moderator variable or complex of moderator variables operating is not readily apparent. Although Vroom and Mann (1960) report that there were virtually no differences due to size of the work group, Dunnette (1963) reports this study as an investigation of group size as a moderator. The study must remain inconclusive with respect to the particular moderating effects operating.

Vroom (1960) and Porter (1962) have studied the moderating effects of situational variables. The effect of the type of situation, motivating or non-motivating on the predictive efficiency of aptitude tests (verbal and non-verbal, arithmetic reasoning) was studied by Vroom (1960). Aptitude tests were found to be correlated positively with job effectiveness in motivating situations and zero or negatively with job effectiveness in non-motivating situations. Porter (1962) studied the effect of size of the organization on the validity of the masculinity-femininity scale of the Strong Vocational Interest Blank. The patterns of validities obtained suggest the hypothesis that masculinity of interests is positively correlated with executive success in small organizations and negatively correlated with executive success in large organizations. The present status of the effort directed toward the investigation of moderating effects of situational variables indicates that these variables are being neglected though they show potential as effective moderators.

In addition to the development of empirical moderator scales by Johnson and Roy (1958), previously reported, other military psychologists have also contributed to the literature on moderator variables. Gaylord and Carroll (1948) investigated the frequency of contact of rater and ratee as a moderator variable and found that efficiency ratings of Army personnel were predicted more effectively by records and grades when contact with rating superior was infrequent, and more effectively by personality characteristics when contact with rating superior was frequent. In a study of Naval personnel, Kipnis (1962) investigated the moderating effect of general intelligence using the Hands Skills Test, a non-cognitive measure which attempts to assess persistence beyond minimum standards. He found differential patterns of validation for the high and low aptitude men. The non-cognitive measure was found to be a valid predictor of school and job performance for the low aptitude men.

Parrish (1959) compared the predictive efficiency of the moderated regression method, a pattern analytic technique, and the more conventional linear multiple regression method. Various aptitude, interest, and personality scales were used as predictors and combat efficiency ratings were used as a criterion. Neither the moderated regression nor the pattern analytic technique was found to be more effective than the linear multiple regression method. He concluded that the results of his study suggest that paper and pencil tests are not particularly effective as moderators. This statement is not supported by other evidence reported herein. Also, Johnson (1960), in a paper on the moderator variable in personnel research, has argued that the empirical evaluation of the usefulness

of moderator variables is most appropriately evaluated by looking at the mean criterion score of the selected group. And, as a corollary to this latter argument, he has further argued that the comparison of validities generated by linear and moderate models should not usually be made in the total sample consisting of combined subgroups, as was done by Parrish (1959). It would appear that reanalysis of the Parrish data is necessary in order to evaluate the predictive efficiency of the various methods studied.

C. Problem Areas in Moderator Research

There is, then, a substantial body of evidence provided by psychologists in educational, industrial, and military settings indicating that certain scales and variables may function as moderators in specific situations. However, even recognizing the existence of moderators and the possibility of differentiating within a group those individuals whose scores are more reliable or more accurate from those whose scores are less reliable or less accurate, a number of questions remain to be answered satisfactorily. For instance, how does one identify a moderator variable? Is it possible to state any general principles about the nature of traits which act as moderators? Can a theoretical foundation be formulated to explain moderator effects? What is the practical significance of moderator variables? Partial answers to these questions are evident in the current literature.

1. Identification of Moderator Variables. It was previously shown that both the rational and empirical approaches have been successfully used to identify moderator variables. However, the method of trying out specific hypotheses is slow and expensive because of the tremendous number of possibilities that can be tried out, and the development of moderator scales through item analysis for each specific situation is certainly no more satisfactory. At the present time, there are no statistics, analogous to multiple regression techniques, to mechanically identify moderators from among a matrix of intercorrelations, and furthermore not enough evidence has accumulated to make it possible to state any general principles about the nature of items or traits that function as moderators. French (1961) did attempt to develop a program for the selection of moderator variables from a large number of predictors, but his indices derived from 3×3 contingency tables, resulting from a pairing of predictors and criterion scores after both had been trichotomized, proved to be unsatisfactory. Unless an effective technique can be developed to identify moderators from massive data, or transsituational moderators can be identified, the utility of moderator variables will be limited. Hence, as Ghiselli (1963b) has cautioned, "It is quite possible that the time and effort required to develop moderators might be more fruitfully spent seeking improvements in reliability and validity of the sort that follow from classic psychometric theory".

2. Explanation of Moderator Effects. Few investigators have given more than cursory attention to the explanation of moderator effects. The bulk of research has been devoted to the problem of substantiating their existence. Of those who have concerned themselves with the problem of explaining these effects, the common notion held is that moderators operate by sorting heterogeneous aggregates of individuals into homogeneous subgroups (Saunders, 1956; Johnson, 1960; Berdie, 1961). As simple as this explanation appears on the surface, it would appear to be deficient since it does not elaborate on the relationship between the moderator, the predictor, and the criterion; and furthermore it fails to consider the special way in which the subgroups are formed. The subgrouping is not simply homogeneous grouping, but it is grouping restricted usually to the extreme scores on the moderator variable. This implies a special mode of subgrouping. Furthermore, this explanation does not do justice to the rather complicated interaction processes which seem to be occurring. However, as Ghiselli (1963) points out, this notion of conceptualizing moderators is useful since it does focus attention on the kinds of differences which exist among individuals who in some given respect are homogeneous thus suggesting types of moderators.

Ghiselli (1963b) explains that in accord with the sorting of heterogeneous aggregates of individuals into homogeneous subgroups, the magnitude and pattern of intercorrelations among variables, and hence the reliability and validity, would vary from group to group. Heterogeneity would be indicated by systematic variation of error from individual to individual whereas homogeneity would be indicated by all individuals having the same error. Such a conceptualization permits retention of the classic psychometric concepts of randomness of errors and the linear combination of variables. This, the psychological structure may vary from group to group since the magnitude of error and the weights carried by the composite may differ for each of the groups, but within a group the error of measurement and of prediction and the relative weights carried by a set of tests in predicting a criterion would be the same for all individuals. This explanation presumes that individuals can be divided into distinct groups, but as previously cited evidence indicated, in actuality, moderators usually distribute individuals along a continuum.

Rather than explain moderator effects in terms of differences between groups, another possible explanation is that the common elements which account for the correlation between two variables differ from individual to individual. What in the first point of view above were considered as classes are now thought of as class intervals. This particular notion appears to receive support from Ghiselli (1963b), Lee (1961), and Johnson (1960), at least where continuous moderator variables are involved.

Ghiselli (1963b) explains that for this conceptualization error of measurement would be taken as being quite small for some individuals and larger for others. Consequently error scores would carry less weight in determining fallible scores for some individuals than for others. Obviously a necessary condition is that individual differences in error scores possess some consistency or reliability over parallel tests. Evidence supporting this is provided by Fiske and Rice (1955), Fiske (1957) and Berdie (1961). Such a position would not require that all variation commonly termed error of measurement is predictable, but only a portion of it. The remainder would still be thought of as being random error. The reliability coefficient, then, would be an average description of precision of measurement.

Also, error of prediction would be taken as being quite small for some individuals and larger for others. The importance of a given trait in determining performance on some criterion is taken to differ among individuals. At one extreme, then, the error prediction is smaller and test validity higher and at the other end, error is larger and test validity lower. Validity coefficients, then, would be average descriptions of predictive accuracy and multiple regression weights indicators of average relative importance of the different predictors. Therefore, with respect to validity, the function of the moderator is to predict for a given individual the weight a test carried in determining criterion performance. As in the case for reliability, it is not necessary that the moderator account for all variation, but only a portion of it. Some of the criterion variance can be due to unreliability and the rest to unmeasured but important traits. The individuals' weights might be unrelated both to their criterion and test scores, or related to one or both. But nothing in this concept indicates what such correlations should be. Perhaps the correlations between the weights and the criterion and the test differ from situation to situation.

A more definitive answer to this question of how to explain moderator effects awaits more theoretically oriented research. If a theoretical foundation were provided, more investigators would be less skeptical about the existence of moderator variables and the future contributions of moderator research.

3. Practical Significance of Moderator Variables. The selected studies cited from the literature substantiate the potential application of the moderator concept in every setting where instruments are used to select, classify, or counsel people. However, even ignoring the problems related to the identification of moderators and their apparent specificity, one might still question their practical value.

The practical value of differentiating individuals with respect to differential error of measurement is clearly evident when administrative decisions concerning borderline cases based on test or measurement devices are to be made. In these borderline cases, it would be very helpful to know whether an individual is measured with a small or large error. Furthermore, a small error of measurement should mean enhancement of validity which Berdie (1961) has empirically demonstrated. In addition, Ghiselli (1963b) has pointed out another practical consideration. In some situations it might be highly desirable to be able to predict the extent of intra-individual variability in performance for planning purposes or to insure smooth flow of work. Therefore, it would be just as important to select individuals whose work did not vary from one period to another as it would be to select individuals whose performance was high. A high performer who was extremely variable could very well reduce the effectiveness of an operation where smooth flow of work was necessary.

The practical value of differentiating individuals with respect to differential error of prediction would be the enhancement of validity resulting from the selection of individuals from those individuals for whom the test is most valid, hence the selection of higher performing people. But this use of moderators might be criticized because it necessitates the elimination of a substantial proportion of the cases from the appraisal procedures, which in turn eliminates even more. This is not a telling criticism, however, since Ghiselli (1956) has demonstrated that if a given percentage of individuals is to be selected and the rest eliminated, selecting that given percentage on the combined basis of their moderator and test scores yields a substantially superior group of individuals than that selected on the basis of test scores alone.

A more powerful way to use the moderator concept and one which permits the assessment of all individuals is to determine which of two or more predictors to use in selection; that is, the moderator is used to determine which of the predictors is suitable for a particular individual being considered. Ghiselli (1960a) has shown that this procedure is effective. The practical implications of this procedure are the increased flexibility of the use of selection instruments and the increased efficiency of the selection procedures. As Dunnette (1963) has so aptly stated, "The selection expert of tomorrow will no longer be attempting to utilize the same procedure for all his selection problems. Instead he will be armed with an array of prediction equations. He will have developed, through research, a wealth of evidence showing the patterns of validities for different linkages in the modified prediction model--for different predictors, candidates, jobs, and criteria. He will be a flexible operator, attentive always to the accumulating information on any given candidate, and ready to apply, at each stage, the tests and procedures shown to be optimal".

REFERENCES

- Abelson, R. P. Sex differences in predictability of grades. *Educ. psychol. Measmt.*, 1952, 12, 638-644.
- Anderson, L. B., & Spencer, Patricia. Personal adjustment and academic predictability among college freshmen. *J. appl. Psychol.*, 1963, 47, 97-100.
- Balma, M. J., Ghiselli, E. E., McCormick, E. J., Primoff, E. S., & Griffin, C. H. The development of processes for indirect or synthetic validation: A symposium. *Personnel Psychol.*, 1959, 12, 395-400.
- Banas, P. A. An investigation of transsituational moderators. Unpublished doctoral dissertation, University of Minnesota, 1964.
- Berdie, R. F. Aptitude, achievement, and personality tests: A longitudinal comparison. *J. appl. Psychol.*, 1955, 39, 103-114.
- Berdie, R. F. Intra-individual variability and predictability. *Educ. psychol. Measmt.*, 1961, 21, 663-676.
- Campbell, D. P., & Trockman, Rachel. A Verification Scale for the Minnesota Vocational Interest Inventory. *J. appl. Psychol.*, 1963, 47, 276-279.
- Court, A. T. Measuring joint causation. *Jour. Amer. Stat. Assoc.*, 1930, 25, 245-254.
- DeSoto, C. B., & Kuethe, J. L. On the relation between two variables. *Educ. psychol. Measmt.*, 1960, 20, 743-749.
- Dunnette, M. D. Validity of interviewer's rating and psychological tests for predicting the job effectiveness of engineers. St. Paul: Minnesota Mining and Manufacturing Company, 1958. (Mimeo)
- Dunnette, M. D. Personnel Management. In *Annual Review of Psychology*, 1962, 13, 285-314.
- Dunnette, M. D. A modified model for test validation and selection research. *J. appl. Psychol.*, 1963, 47, 317-323.
- Dunnette, M. D., & Kirchner, W. K. Validation of psychological tests in industry. *Personnel Admin.*, 1958, 21, 20-27.
- Dunnette, M. D., & Kirchner, W. K. Psychological test differences between industrial salesmen and retail salesmen. *J. appl. Psychol.*, 1960, 44, 121-125.

- Eells, K. How effective is differential prediction in three types of college curricula? *Educ. Psychol. Measmt.*, 1961, 21, 459-471.
- England, G. W. The interest factor in undergraduate engineering achievement. *Personnel Guid.*, 1960, 39, 401-405.
- Filbeck, R., & Callis, R. A verification Scale for the Strong Vocational Interest Blank. *J. appl. Psychol.*, 1961, 45, 318-342.
- Fiske, D. W. Homogeneity and variation in measuring personality. *Amer. Psychol.*, 1963, 18, 643-652.
- Fiske, D. W. & Rice, L. Intra-individual response variability. *Psychol. Bull.*, 1955, 52, 217-250.
- Frederiksen, N., & Gilbert, A. C. Replication of a study of differential predictability. *Educ. psychol. Measmt.*, 1960, 20, 759-767.
- Frederiksen, N., & Malville, S. D. Differential predictability in the use of test scores. *Educ. psychol. Measmt.*, 1954, 14, 647-656.
- French, J. W. A Machine Search for Moderator Variables in Massive Data. Princeton, N. J.: Educational Testing Service, 1961, 19 pp.
- Gallese, A. J. Personality characteristics and academic achievement in school of engineering students. *Dissertation Abstracts*, 1958, 19, 3022-23.
- Gaylord, R. H., & Carroll, J. B. A general approach to the problem of the population control variable (abstract). *Amer. Psychologist*, 1948, 310.
- Ghiselli, E. E. The Measurement of Occupational Aptitude. Berkley & Los Angeles, Calif.: University of California Press, 1955.
- Ghiselli, E. E. Differentiation of individuals in terms of their predictability. *J. appl. Psychol.*, 1956, 40, 374-377.
- Ghiselli, E. E. Differentiation of tests in terms of the accuracy with which they predict for a given individual. *Educ. psychol. Measmt.*, 1960, 20, 675-684. (a)
- Ghiselli, E. E. The prediction of predictability. *Educ. psychol. Measmt.*, 1960, 20, 3-8. (b)
- Ghiselli, E. E. Models of True Scores: Così e se vi para. Donald G. Paterson Memorial Lecture, presented at the University of Minnesota, Minneapolis, Minnesota, March, 1963. (a)

- Ghiselli, E. E. Moderating effects and differential reliability and validity. *J. appl. Psychol.*, 1963, 47, 81-86. (b)
- Goodstein, L., & Heilbrun, A. B. Prediction of college achievement from the Edwards Personal Preference Schedule at three levels of intellectual ability. *J. appl. Psychol.*, 1962, 46, 317-320.
- Gordon, L. V. *Gordon Personal Profile: Manual*. New York: Harcourt, Brace, & World, 1963. (a)
- Gordon, L. V. *Gordon Personal Inventory: Manual*. New York: Harcourt, Brace, & World, 1963. (b)
- Grooms, R. R., & Endler, N. S. The effect of anxiety on academic achievement. *J. educ. Psychol.*, 1960, 51, 299-304.
- Guetzkow, H., & Forehand, G. A. A Research Strategy for Partial Knowledge Useful in the Selection of Executives. In Taguirl, R. (Ed.), *Research Needs in Executive Selection*. Boston: Harvard Graduate School of Business Administration, 1961.
- Holland, J. L. The prediction of college grades from the California Psychological Inventory and Scholastic Aptitude Test. *J. educ. Psychol.*, 1960, 51, 245-254.
- Holland, J. L., & Astin, A. W. The prediction of the academic, artistic, scientific, and social achievement of undergraduates of superior scholastic aptitude. *J. educ. Psychol.*, 1962, 53, 132-143.
- Holtzman, W. H., Brown, W. F., & Farquhar, W. G. The survey of study habits and attitudes: A new instrument for the prediction of academic success. *Educ. psychol. Measmt.*, 1954, 14, 726-732.
- Horst, P. A. A Technique for the Development of a Differential Prediction Battery. *Psychol. Monogr.* 48, 1954, No. 9 (Whole No. 380).
- Horst, P. A. A Technique for the Development of a Multiple Absolute Prediction Battery. *Psychol. Monogr.* 69, 1955, No. 5 (Whole No. 390).
- Horst, P. A. *Relationship between Preadmission Variables and Success in College*. Seattle, Wash.: Univ. of Washington, Division of Counseling and Testing Services, 1959.
- Hoyt, D. F., & Norman, W. T. Adjustment and academic predictability. *J. counsel. Psychol.*, 1954, 1, 96-99.

- Hull, C. L. Variability in amount of different traits possessed by the individual. J. educ. Psychol., 1927, 28, 97-104.
- Hull, C. L. Aptitude Testing. Yonkers, N. Y.: World Book, 1928.
- Johnson, C. D. The population control variable or moderator variable in personnel research. In Tri-Service Conference on Selection Research, 125-134. Washington, D. C.: Office of Naval Research, 1960, 273 pp.
- Johnson, C. D., & Roy, H. L. The use of a "Carelessness" key to refine analysis samples. Washington, D. C.: Personnel Research Branch, AGO, Department of Army, 1958.
- Kahneman, D., & Ghiselli, E. E. Validity and non-linear heteroscedastic models. Educ. psychol. Measmt., 1962, 15, 1-11.
- Kipnis, D. A noncognitive correlate of performance among lower aptitude men. J. appl. Psychol., 1962, 46, 76-80.
- Kuder, G. F. Manual for the Kuder Preference Record, Form C. Chicago: Science Research Associates, 1956.
- Lee, Marilyn C. Interactions, configurations, and non-additive models, Educ. psychol. Measmt., 1961, 21, 797-805.
- Malnig, L. R. Differential Predictability of Academic Achievement from Aptitude Tests Administered under Stress to Subjects with Varying Amounts of Anxiety. Unpublished Ph. D. thesis, New York University, 1959.
- Marks, M. R. Derivation of moderator variables: theoretically and from data. Paper presented at meeting of American Psychological Association; Los Angeles, California, September 1964.
- McArthur, C. Long term validity of the Strong Interest Test in two subcultures. J. appl. Psychol., 1954, 38, 346-353.
- Meehl, P. E., & Hathaway, S. R. The K factor as a suppressor variable in the MMPI. J. appl. Psychol., 1946, 30, 525-564.
- Neidt, C. O., & Malloy, J. P. A technique for keying items of an inventory to be added to an existing test battery. J. appl. Psychol., 1954, 38, 308-312.
- Parrish, J. A. A Study of Two Non-linear Methods of Combining Predictor Tests. PRB Tech. Note 103. Washington, D. C.: Personnel Research Branch, AGO, Dept. of Army, 1959, 21 pp.

- Porter, A. Effect of organization size on validity of masculinity-femininity score. *J. appl. Psychol.*, 1962, 46, 228-229.
- Primoff, E. S. Test selection by job analysis. Washington, D. C.: United States Civil Service Commission, Test Development Section, 1955.
- Rundquist, E. A. Sex, intelligence, and school marks. *Sch. Soc.*, 1941, 53, 452-456.
- Saunders, D. R. Moderator variables in prediction. *Educ. psychol. Measmt.*, 1956, 16, 209-222.
- Scates, D. E. Selective admission and retention of college students at the University of Chicago. Ph.D. thesis, School of Education, University of Chicago, 1926.
- Seashore, H. G. Women are More Predictable than Men. Presidential Address, Division of Counseling Psychology, presented at 69th annual convention of the American Psychological Association, September 1, 1961, New York.
- Stagner, R. The relation of personality to academic aptitude and achievement. *J. educ. Res.*, 1933, 26, 648-660.
- Toops, H. A. The use of addends in experimental control, social census, and managerial research. *Psychol. Bull.*, 1948, 45, 41-74.
- Toops, H. A. A research utopia in industrial psychology. *Personnel Psychol.*, 1959, 12, 189-225.
- Uhlinger, Carolyn A., & Stephens, M. W. Relation of achievement motivation to academic achievement in students of superior ability. *J. educ. Psychol.*, 1960, 51, 259-266.
- Vroom, V. H. Some Personality Determinants of the Effects of Participation. Englewood Cliffs, New Jersey: Prentice-Hall, 1960.
- Vroom, V. H., & Mann, F. C. Leader authoritarianism and employee attitudes. *Personnel Psychol.*, 1960, 13, 125-140.
- Wagner, M. E., & Strabel, E. Homogeneous grouping as a means of improving prediction of academic performance. *J. appl. Psychol.*, 1935, 19, 426-446.
- Wallace, S. R., Jr. & Twichell, C. M. Managerial procedures and test validities. *Personnel Psychol.*, 1949, 2, 277-292.